EMBL-EBI

Update: MIRIAM Registry and SBO

Nick Juty, EMBL-EBI

3rd Sept, 2011

COMBINE, Heidelberg, Germany 3-7th Sept, 2011
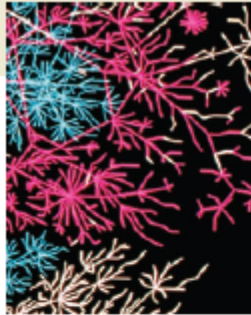
EMBL-EBI

MIRIAM Registry

- MIRIAM Guidelines ..

- MIRIAM Registry

    - content

    - URIs (URN form), example

    - Summary/current developments

SBO

- Purpose and origins

- Term information stored

- Summary of changes since Edinburgh

http://biomodels.net/miriam

_computational BIOLOGY

PERSPECTIVE

Minimum information requested in the annotation of biochemical models (MIRIAM)

Nicolas Le Novère[1,15], Andrew Finney[2,15], Michael Hucka[3], Upinder S Bhalla[4], Fabien Campagne[5], Julio Collado-Vides[6], Edmund J Crampin[7], Matt Halstead[7], Edda Klipp[8], Pedro Mendes[9], Poul Nielsen[7], Herbert Sauro[10], Bruce Shapiro[11], Jacky L Snoep[12], Hugh D Spence[13] & Barry L Wanner[14]

Most of the published quantitative models in biology are lost for the community because they are either not made available or they are insufficiently characterized to allow them to be reused. The lack of a standard description format, lack of stringent reviewing and authors' carelessness are the main causes for incomplete model descriptions. With today's increased interest in detailed biochemical models, it is necessary to define a minimum quality standard for the encoding of those models. We propose a set of rules for curating quantitative models of biological systems. These rules define procedures for encoding and annotating models represented in machine-readable form. We believe their

During the genomic era we have witnessed a vast increase in availability of large amounts of quantitative data. This is motivating a shift in the focus of molecular and cellular research from qualitative descriptions of biochemical interactions towards the quantification of such interactions and their dynamics. One of the tenets of systems biology is the use of quantitative models (see Box 1 for definitions) as a mechanism for capturing precise hypotheses and making predictions[1,2]. Many specialized models exist that attempt to explain aspects of the cellular machinery. However, as has happened with other types of biological information, such as sequences, macromolecular structures or

Models must :

- be encoded in a public machine-readable format, standard compliant

- be named and clearly linked to a single reference description

- distribution terms

- contain creator's contact details

- reflect the structure of the biological processes described in the reference paper (list of reactions etc.)

- be instantiable in a simulation (possess initial conditions etc.)

- be able to reproduce the results given in the reference paper

- annotation to unambiguously identify each model constituent ...

Annotation specification:

- unambiguously relate component to external resource

- be encoded in a defined triplet format

  - as a URI

  - identifier unique for the data type

  - optional use of qualifiers to refine relationship

  - use a defined set of data types – community agreed

    - approved data types stored in MIRIAM Registry

- A data type is a set of data within which all data points refer to comparable entities, in terms of the 'properties' they can/do possess

- Individual data points (entities/records) share a common identifier scheme which can be used to distinguish them

- Data of the same 'kind'
  - UniProt – protein data
  - Chebi – small molecules
  - PubMed – publication records

- **Open access**

Anybody can access any public data without restriction (no commercial licence; no login page etc.)

- **Atomicity**

The granularity of the data distributed has to be appropriately selected (A database of "reactions" distributes reactions and not pathways) and consistent (e.g. classes or instances but not classes AND instances)
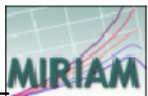
- **Identifier**

An atomic data is associated to a unique and perennial identifier

- **Community recognition**

The resource has to be "recognised" by the corresponding experimental community, be reasonably supported etc

MIRIAM Registry

EBI > Groups > Computational Neurobiology > Research > MIRIAM Registry

MIRIAM Registry

MIRIAM Registry are a set of online services created in support of MIRIAM, a set of guidelines for the annotation and curation of computational models.

The core of *MIRIAM Registry* is a catalogue of data types (namespaces corresponding to controlled vocabularies or databases), their URIs and the corresponding physical URLs or resources. Access to this data is made available via exports (XML) and Web Services (SOAP).

MIRIAM Registry is developed and maintained under the BioModels.net initiative, and are free for use by all.
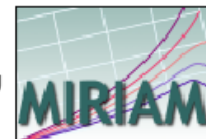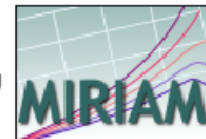
Quick links

Browse
- by data type name
- by tags

Web Services
- services available
- usage of the services
- online demonstration

Search
- generic search

Exports
- XML

Registry

MIRIAM Registry is composed of four components: a database, some Web Services, a Java library and this web application.

Sidebar:
- Browse
- Search
- Tags
- Query services
- Submit new
- Export
- Curator Sign in
- Web Services
- Documents
  - MIRIAM Guidelines
  - FAQ
  - Documentation
  - Who's using MIRIAM?
  - Identification systems
  - News
  - BioModels.net Qualifiers
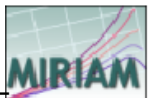- MIRIAM on SourceForge

COMBINE, Heidelberg, Germany 3-7th Sept, 2011

http://www.ebi.ac.uk/miriam/

EBL-EBI

MIRIAM

- Browse
- Search
- Tags
- Query services
- Submit new
- Export
- Curator Sign in

Web Services

Documents
- MIRIAM Guidelines
- FAQ
- Documentation
- Who's using MIRIAM?
- Identification systems
- News
- BioModels.net Qualifiers

MIRIAM on SourceForge

- Support
- Contact

EBI > Groups > Computational Neurobiology > Research > MIRIAM Registry

Browse data types: *recently updated*

Recently updated | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

Name	Namespace	Definition
UniProt	uniprot	UniProt (Universal Protein Resource) is the world's most comprehensive catalog of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR.
PubChem-substance	pubchem.substance	PubChem provides information on the biological activities of small molecules. It is a component of NIH's Molecular Libraries Roadmap Initiative. PubChem Substance archives chemical substance records.
EDAM Ontology	edam	EDAM is an ontology of general bioinformatics concepts, including topics, data types, formats, identifiers and operations. EDAM provides a controlled vocabulary for the description, in semantic terms, of things such as: web services (e.g. WSDL files), applications, tool collections and packages, work-benches and workflow software, databases and ontologies, XSD data schema and data objects, data syntax and file formats, web portals and pages, resource catalogues and documents (such as scientific publications).
Ontology for Biomedical Investigations	obo.obi	The Ontology for Biomedical Investigations (OBI) project is developing an integrated ontology for the description of biological and clinical investigations. The ontology will represent the design of an investigation, the protocols and instrumentation used, the material used, the data generated and the type analysis performed on it. Currently OBI is being built under the Basic Formal Ontology (BFO).
MACiE	macie	MACiE (Mechanism, Annotation and Classification in Enzymes) is a database of enzyme reaction mechanisms. Each entry in MACiE consists of an overall reaction describing the chemical compounds involved, as well as the species name in which the reaction occurs. The individual reaction stages for each overall reaction are listed with mechanisms, alternative mechanisms, and amino acids involved.
FMA	obo.fma	The Foundational Model of Anatomy Ontology (FMA) is a biomedical informatics ontology. It is concerned with the representation of classes or types and relationships necessary for the symbolic representation of the phenotypic structure of the human body. Specifically, the FMA is a domain ontology that represents a coherent body of explicit declarative knowledge about human anatomy.
OMIA	omia	Online Mendelian Inheritance in Animals is a a database of genes, inherited disorders and traits in animal species (other than human and mouse).
MIRIAM Registry data type	miriam.datatype	MIRIAM Registry is an online resource created to catalogue data types (Gene Ontology, Taxonomy or PubMed are some examples), their URIs and the corresponding resources (physical locations), whether these are controlled vocabularies or databases.
MIRIAM Registry resource	miriam.resource	MIRIAM Registry is an online resource created to catalogue data types (Gene Ontology, Taxonomy or PubMed are some examples), their URIs and the corresponding resources (or physical locations), whether these are controlled vocabularies or databases.
Anatomical Therapeutic Chemical	atc	The Anatomical Therapeutic Chemical (ATC) classification system, divides active substances into different groups according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties. Drugs are classified in groups at five different levels; Drugs are divided into fourteen main groups (1st level), with pharmacological/therapeutic subgroups (2nd level). The 3rd and 4th levels are chemical/pharmacological/therapeutic subgroups and the 5th level is the chemical substance. The Anatomical Therapeutic Chemical (ATC) classification system and the Defined Daily Dose (DDD) is a tool for exchanging and comparing data on drug use at international,

Data type: *Enzyme Nomenclature*

General | Tags | Example Usage | Web Services

General information about the data type

Name

Identifier	MIR:00000004
Name	Enzyme Nomenclature
Synonyms	EC code
	Enzyme Classification
	EC

URIs

MIRIAM URN	urn:miriam:ec-code
Deprecated	http://www.ec-code.org/
	urn:lsid:ec-code.org
	http://www.ebi.ac.uk/IntEnz/

Information

Definition	The Enzyme Classification contains the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzyme-catalysed reactions.
Identifier Pattern	^\d+\.-\.-\.-|\d+\.\d+\.-\.-|\d+\.\d+\.\d+\.-|\d+\.\d+\.\d+\.\d+$

Physical Locations

Resource		
Resource MIR:00100002	Access URL	http://www.genome.jp/dbget-bin/www_bget?ec:$id [Example: 1.1.1.1]
	Website	http://www.genome.jp/dbget-bin/www_bfind?enzyme
	Description	KEGG Ligand Database for Enzyme Nomenclature
	Institution	Kyoto University Bioinformatics Center, Japan
Resource MIR:00100003	Access URL	http://us.expasy.org/cgi-bin/nicezyme.pl?$id [Example: 1.1.1.1]
	Website	http://us.expasy.org/enzyme/
	Description	Enzyme nomenclature database, ExPASy (Expert Protein Analysis System)
	Institution	Swiss Institute of Bioinformatics, Switzerland
Resource MIR:00100001	Access URL	http://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=$id [Example: 1.1.1.1]
	Website	http://www.ebi.ac.uk/intenz/
	Description	IntEnZ (Integrated relational Enzyme database)
	Institution	European Bioinformatics Institute, United Kingdom

References

URL(s)	http://www.chem.qmul.ac.uk/iubmb/enzyme/
	http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-view+MedlineFull+[medline-PMID:10812475]

Miscellaneous

Date of creation	2006-08-14 19:38:06 GMT
Date of last modification	2010-12-09 08:45:47 GMT

COMBINE, Heidelberg, Germany 3-7th Sept, 2011

| General | Tags | Example Usage | Web Services |

General information about the data type

Name	
Identifier	MIR:00000004
Name	Enzyme Nomenclature
Synonyms	EC code
	Enzyme Classification
	EC

URIs	
MIRIAM URN	urn:miriam:ec-code
Deprecated	http://www.ec-code.org/
	urn:lsid:ec-code.org
	http://www.ebi.ac.uk/IntEnz/

Information	
Definition	The Enzyme Classification contains the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzyme-catalysed reactions.
Identifier Pattern	^\d+\.-\.-\.-|\d+\.\d+\.-\.-|\d+\.\d+\.\d+\.-|\d+\.\d+\.\d+\.\d+$

Physical Locations

Resource MIR:00100002	Access URL	http://www.genome.jp/dbget-bin/www_bget?ec:$id [Example: 1.1.1.1 ☒]
	Website	http://www.genome.jp/dbget-bin/www_bfind?enzyme
	Description	KEGG Ligand Database for Enzyme Nomenclature
	Institution	Kyoto University Bioinformatics Center, Japan
Resource MIR:00100003	Access URL	http://us.expasy.org/cgi-bin/nicezyme.pl?$id [Example: 1.1.1.1 ☒]
	Website	http://us.expasy.org/enzyme/
	Description	Enzyme nomenclature database, ExPASy (Expert Protein Analysis System)
	Institution	Swiss Institute of Bioinformatics, Switzerland
Resource MIR:00100001	Access URL	http://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=$id [Example: 1.1.1.1 ☒]
	Website	http://www.ebi.ac.uk/intenz/
	Description	IntEnZ (Integrated relational Enzyme database)
	Institution	European Bioinformatics Institute, United Kingdom

Resources

References	
URL(s)	http://www.chem.qmul.ac.uk/iubmb/enzyme/
	⊞ http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-view+MedlineFull+[medline-PMID:10812475]

| Miscellaneous	
Date of creation	2006-08-14 19:38:06 GMT
Date of last modification	2010-12-09 08:45:47 GMT

EMBL-EBI

Resource: MIR:00100050

General information about the resource: **The FlyBase Database**   (associated with the data type: FlyBase).

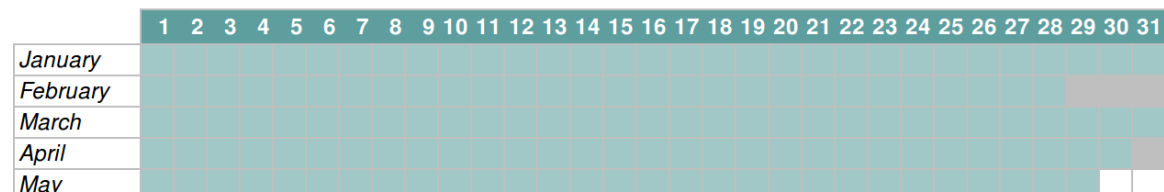| Health statistics	
Last known state	up
Last check	2011-03-29 06:37:48
Uptime ratio	100% (763 checks)
Downtime ratio	0% (0 checks)
Unknown ratio	0% (0 checks)
URL used	http://www.flybase.org/reports/FBgn0011293.html

Health history

Full record of the health checks performed on this resource.

2011
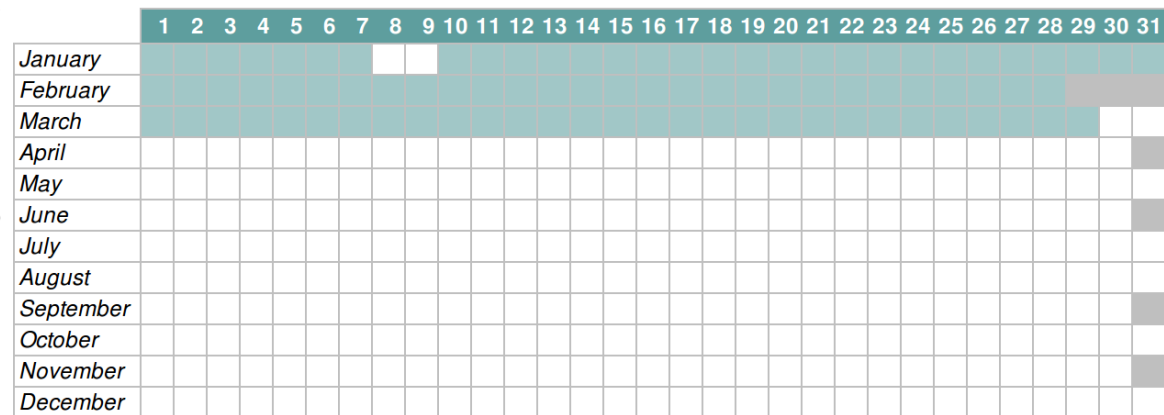
Legend
- up and responsive
- probably up
- resource down
- state unknown
- no data available

2010

Resource: MIR:00100042

WARNING! This resource has been deprecated!
Please use the other resources providing the same dataset.

General information about the resource: **WormBase Mirror at Marseille-Nice Genopole** (associated with the data type: WormBase).

Health statistics	
Last known state	down
Last check	2010-09-15 09:11:48
Uptime ratio	43% (245 checks)
Downtime ratio	56% (318 checks)
Unknown ratio	0% (0 checks)
URL used	http://crfb-3.univ-mrs.fr/db/gene/gene?name=WBGene00000001;class=Gene

Health history

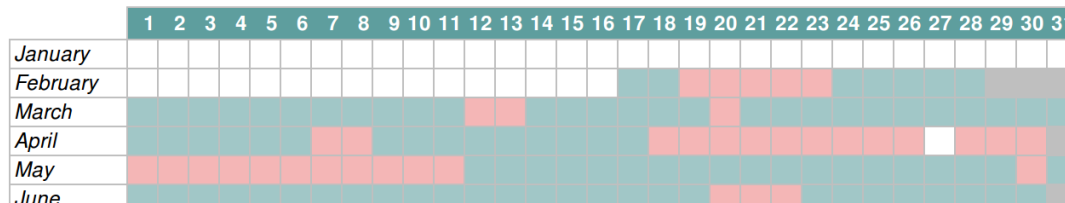Full record of the health checks performed on this resource.

2010
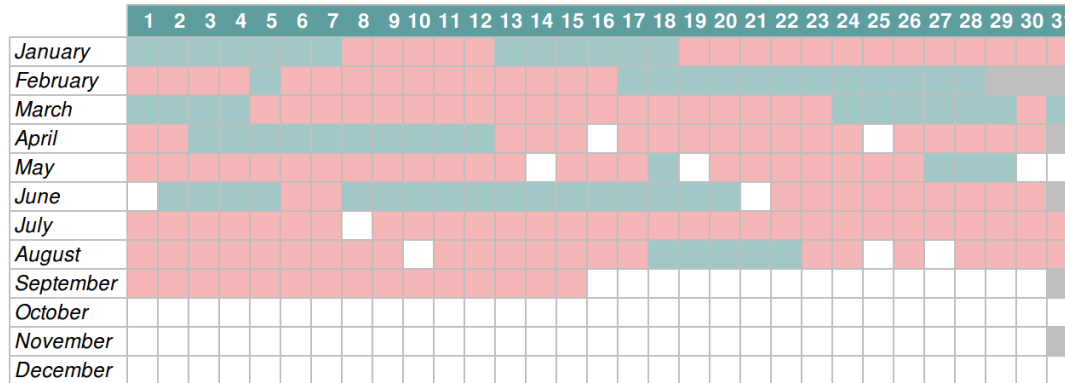
Legend
- up and responsive
- probably up
- resource down
- state unknown
- no data available

2009

Add a data type

Please fill this form in order to submit a new data type to MIRIAM Resources. Alternatively, you can contact us with your query.

The new data type will not be directly publicly available after you pressed the **Submit** button. A curator will first check if it complies with the terms and conditions of MIRIAM Resources before publishing it.

Help

You can display all help bubbles by clicking on: Displays all the help messages, or hide them: Hides all the help messages.

Moreover, you can display the individual help by clicking on the button: located in the title of each section.

Add a new data type

First you need to enter the name of the data type you want to add to the database. After you can add as much synonyms as you want.

Name and synonyms

Primary name:

[Add a synonym]

Here is some information about the data type: definition and regular expression (*i.e.* pattern for identifiers of elements, following the PERL style).

Definition and pattern

Definition: Enter definition here...

Identifier pattern: Enter Identifier pattern here...

COMBINE, Heidelberg, Germany 3-7th Sept, 2011

EMBL-EBI

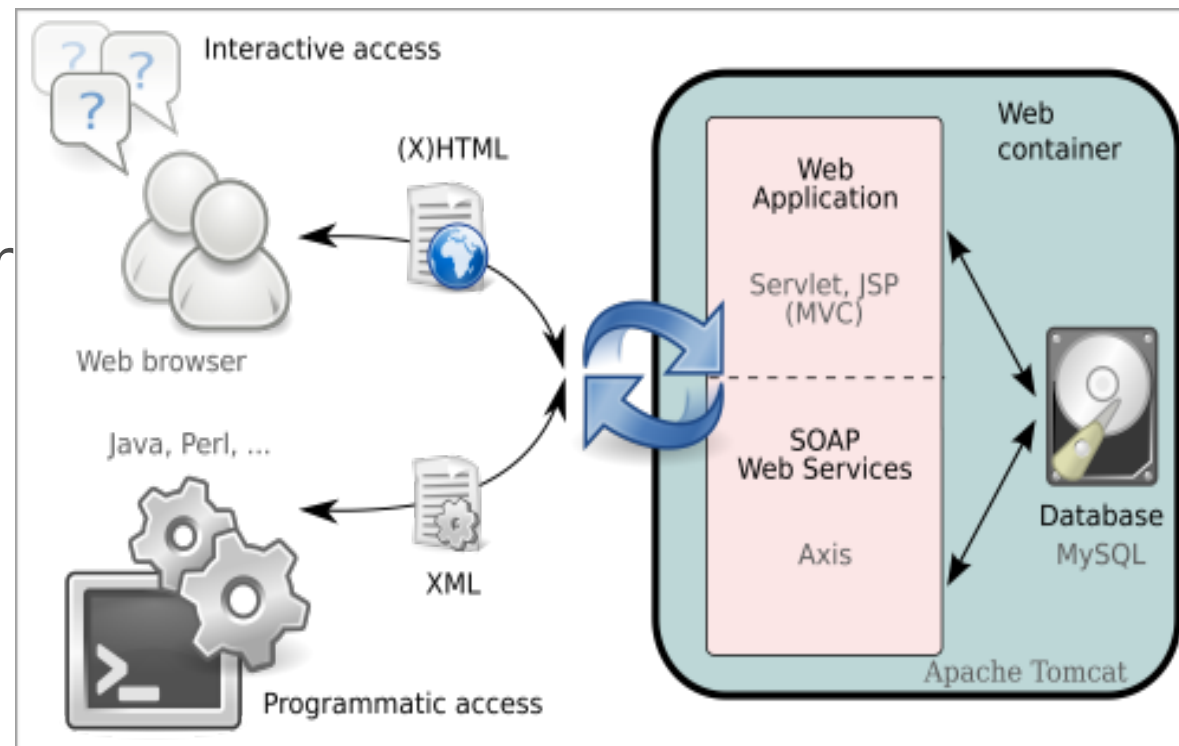## MIRIAM Database

data type catalogue

MIRIAM Web Services

programming interface for querying database

MIRIAM Library

ready to use Java code

MIRIAM Web Site

- web page for browsing and querying
- allows new submission (curated)

Laibe and Le Novère.
MIRIAM Resources: tools to generate and resolve robust cross-references in Systems Biology.
BMC Systems Biology, 2007

EMBL-EBI

Web Services Demonstration

Here is a Web interface which allows you to perform some queries on MIRIAM Database.

These queries are an example of what MIRIAM Web Services can provide you.

For more information, please refer to the complete list of available queries.

Query

Choose one action from the list: | get the MIRIAM URI of an element | ⬍ | Go!

Get the MIRIAM URI of an element or entity:

getURI

data type name | pubmed

element id | 18078503

Search

Answer

urn:miriam:pubmed:18078503

Web Services Demonstration

Here is a Web interface which allows you to perform some queries on MIRIAM Database.

These queries are an example of what MIRIAM Web Services can provide you.

For more information, please refer to the complete list of available queries.

Query

Choose one action from the list: get links to access an element ▼   **Go!**

Get links to access an element:

getLocations

MIRIAM URI urn:miriam:pubmed:18078503

Search

Answer

- http://www.ncbi.nlm.nih.gov/pubmed/18078503
- http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-view+MedlineFull+[medline-PMID:18078503]
- http://www.ebi.ac.uk/citexplore/citationDetails.do?dataSource=MED&externalId=18078503
- http://www.hubmed.org/display.cgi?uids=18078503

Data type
identifier

Data-set
Identifier

Annotation
qualifier

↑

↑

↑

scheme
identifier

text string

optional text
string

urn:miriam:pubmed:16333295 (MIRIAM publication)

urn:miriam:chembl.compound:chembl113 (Caffeine)

- Qualifiers are a Controlled vocabulary (CV)

- CV established by the community

- http://biomodels.net/qualifiers/

```
<species id="Ca_calmodulin" metaid="cacam">
  <annotation>
    <rdf:RDF
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
      <rdf:Description rdf:about="#cacam">
        <bqbiol:hasPart>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:uniprot:P62158"/>
            <rdf:li rdf:resource="urn:miriam:obo.chebi:CHEBI%3A29108"/>
          </rdf:Bag>
        </bqbiol:hasPart>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</species>
```

cacam hasPart P62158 and CHEBI:29108

model component   relationship   external resource(s)

- Introduction of a **URL scheme**, in addition to the URN namespace

- In discussions with *Bio2RDF*

- Collaboration with *BioDbCore*

- More structured tags

- Growth in content/coverage

- Limited support 'branch'

- ...

The Systems Biology Ontology

http://biomodels.net/sbo

EMBL-EBI

The Open Biomedical Ontologies

| Ontologies | Resources | Participate | About |

The OBO Foundry is a collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. The groups developing ontologies who have expressed an interest in this goal are listed below, followed by other relevant efforts in this domain.

In addition to a listing of OBO ontologies, this site also provides a statement of the OBO Foundry principles, discussion fora, technical infrastructure, and other services to facilitate ontology development. We welcome feedback and encourage participation.

Click any column header to sort the table by that column. The 🐾s link to the term request trackers for the listed ontologies.

OBO Foundry candidate ontologies

Title	Domain	Prefix	File	Last changed
Amphibian gross anatomy	anatomy	AAO	amphibian_anatomy.obo	2008/06/19
Amphibian taxonomy	anatomy	ATO	amphibian_taxonomy.obo	
Ascomycete phenotype ontology	phenotype	APO	ascomycete_phenotype.obo	2009/07/10
Biological process	biological process	GO	gene_ontology_edit.obo 🐾	2009/08/21
C. elegans development	anatomy	WBls	worm_development.obo	2008/01/31
C. elegans gross anatomy	anatomy	WBbt	WBbt.obo 🐾	2009/08/19
C. elegans phenotype	phenotype	WBPhenotype	worm_phenotype.obo	2009/08/19
Cell type	anatomy	CL	cell.obo 🐾	2008/12/08
Suggested Ontology for Pharmacogenomics	health	SOPHARM	sopharm	
Systems Biology	biochemistry	SBO	SBO_OBO.obo 🐾	
Teleost anatomy and development	anatomy	TAO	teleost_anatomy.obo 🐾	2009/06/11

The Systems Biology Ontology
http://biomodels.net/sbo

- A navigable taxonomic structure of terms that has 'parents', 'children'

- Provide a strictly defined relational vocabulary of terms for use in Computational Biology

- Describe model components and how they are intended to be used

- 7 orthogonal vocabularies containing domain knowledge:

  - type of entity (compartment / macromolecule)

  - roles of entities (modifier / product)

  - how they interact (reaction / transport)

  - ...

COMBINE, Heidelberg, Germany 3-7th Sept, 2011

Semantic layer:

- conversion to semantically enriched computing formats (such as BioPAX)

- link between models encoded in SBML and graphical notations (such as SBGN)

- translation of models between *continuous deterministic frameworks* and *discrete stochastic framework*

- merging and integration of models

https://sourceforge.net/projects/sbo

Systems Biology Ontology

Share ›   Monitor ›   Watch ›

Summary   Files   Reviews   Support   Develop   Tracker   Forums   Code

Add new   **Browse**   Reporting   Admin

Tracker: Term request

List of suggested SBO term creations or modifications.

Search: [   ]   **Search**   Advanced

Options   RSS

Page: **Page 1**   **Next →**

1 - 25 of 81 Results - Display [25 ▼]

ID	Summary	Status	Opened	Assignee	Submitter	Priority
Assignee: [Any ▼] Status: [Any ▼] Category: [Any ▼] Group: [Any ▼] Submitter: [ ] Keyword: [ ] Artifact ID:						
[ ] **Filter** **Reset** Permalink						
☐ 3356372	Ion channel terms (SBPAX/SGMP)	Open	2011-07-06	njuty	njuty	6
☐ 3325167	mass action rate laws for essential stimulators	Closed	2011-06-23	njuty	njuty	5
☐ 3314182	sbo terms for reaction rates	Open	2011-06-09	nobody	nobody	5
☐ 3308330	SBO:0000061 has wrong rate constant	Closed	2011-05-27	njuty	njuty	5
☐ 3301506	specific activity	Closed	2011-05-13	njuty	njuty	4
☐ 3293425	Goldman & Nernst Equations	Open	2011-04-27	njuty	luen	5
☐ 3286595	Gibbs free energy of reaction	Closed	2011-04-14	njuty	nobody	6

COMBINE, Heidelberg, Germany 3-7th Sept, 2011

Problem: 'substrate' (participant role) = amount? concentration?

SBML component

SBO branch

<speciesReference>

substrate
product

participant role

quantitative parameter

<kineticLaw>

rate law

MathML

<bvar>
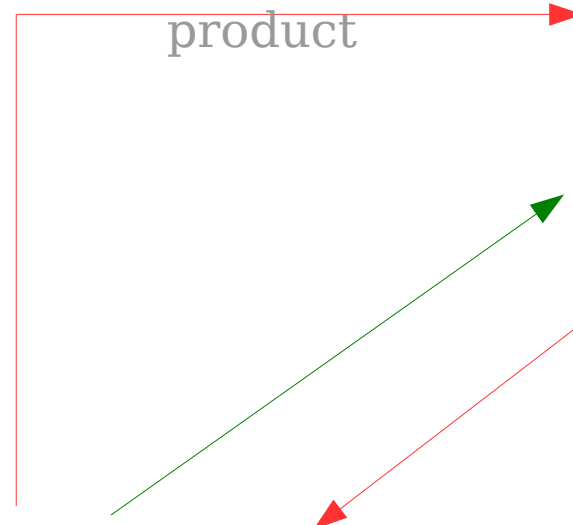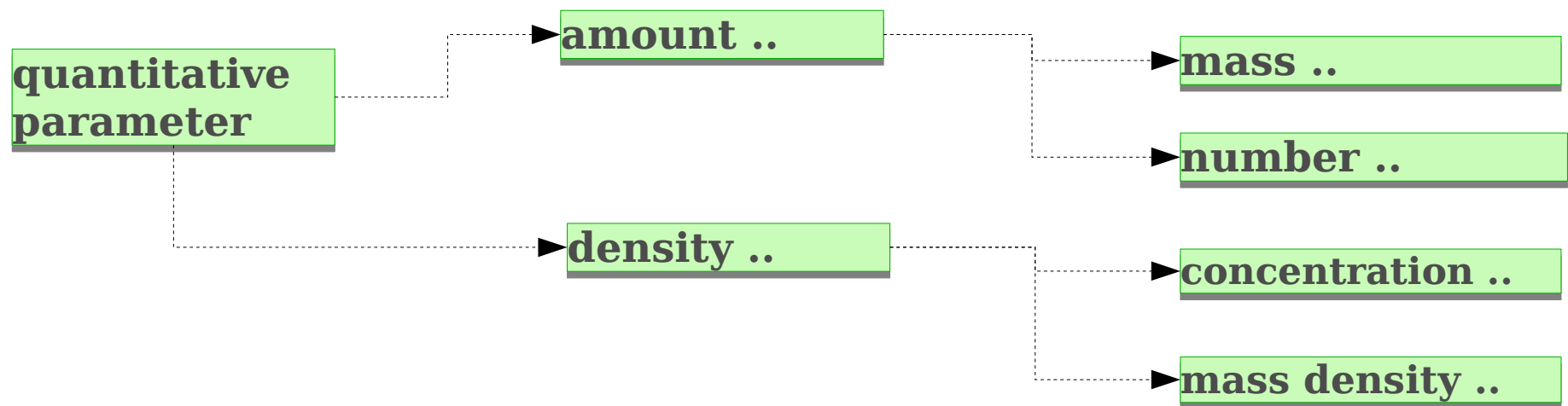SBO:0000015"> substrate <..
SBO:0000011"> product   <..

COMBINE, Heidelberg, Germany 3-7th Sept, 2011

- Initial changes made on SBO 'demo'

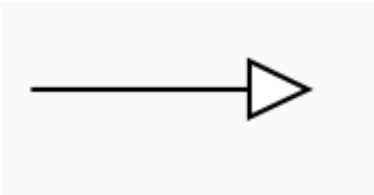- Introduce 'quantitative parameter's for 'participant role' terms

- Modify <bvar> links in all MathML - redirect to quantitiative parameter branches

- Request feedback for selected users (validate)

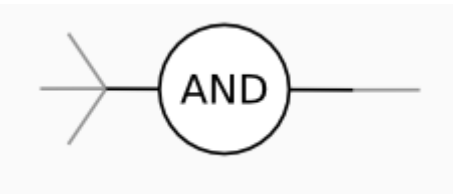- Move changes to SBO 'main' (implement → live version)

COMBINE, Heidelberg, Germany 3-7th Sept, 2011

EMBL-EBI
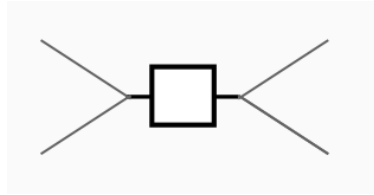
SBGN	<--------------->	SBO
	Process (PD)	SBO:0000375
	Stimulation (ER)	SBO:0000170
	And (AF)	SBO:0000173

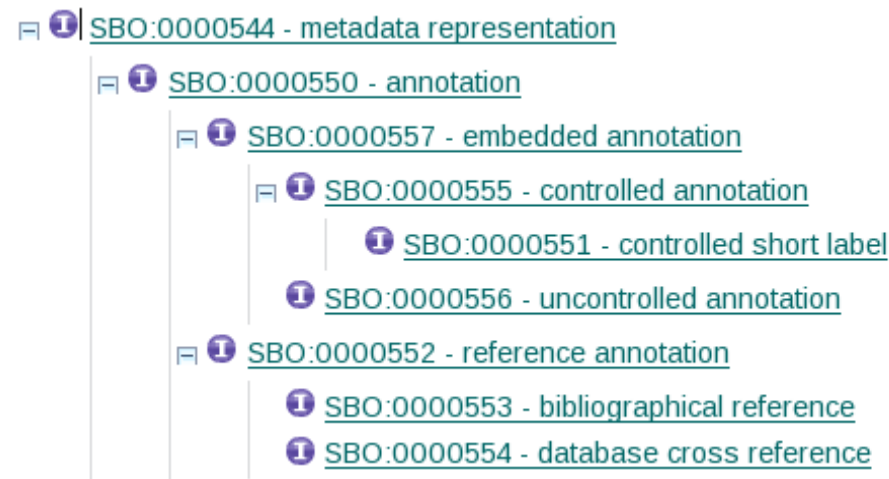**SBGN**   <--------------->   **SBO**

Annotation
(ER)

?

- 7th branch: Annotation, cross-references, metadata

Acknowledgements

SBO + MIRIAM

- Mélanie Courtot
- Camille Laibe
- Nicolas Le Novère
- Lukas Endler

SBML team

- Michael Hucka
- Sarah Keating
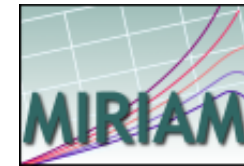
BioModels Database
developers and curators

The Systems Biology community for their contributions, software support and their comments.

```xml
<species id="Ca_calmodulin" metaid="cacam">
  <annotation>
    <rdf:RDF
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
      <rdf:Description rdf:about="#cacam">
        <bqbiol:hasPart>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:uniprot:P62158"/>
            <rdf:li rdf:resource="urn:miriam:obo.chebi:CHEBI%3A29108"/>
          </rdf:Bag>
        </bqbiol:hasPart>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</species>
```
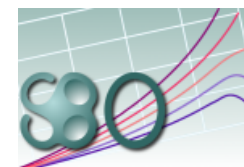
```
<species id="Ca_calmodulin" metaid="cacam">
  <annotation>
    <rdf:RDF
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
      <rdf:Description rdf:about="#cacam">
        <bqbiol:hasPart>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:uniprot:P62158"/>
            <rdf:li rdf:resource="urn:miriam:obo.chebi:CHEBI%3A29108"/>
          </rdf:Bag>
        </bqbiol:hasPart>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</species>
```

cacam hasPart P62158 and CHEBI:29108